THE DETERMINANTS OF BASE PAY AND THE ROLE OF RACE
IN MAJOR LEAGUE SOCCER: EVIDENCE FROM THE 2007
LEAGUE SEASON

By

Barry Reilly

(University of Sussex)

&

Robert Witt

(University of Surrey)

# The Determinants of Base Pay and the Role of Race in Major League Soccer: Evidence from the 2007 League Season

**Barry Reilly**

**and**

**Robert Witt**

## Abstract

*This paper examines pay determination in the labor market of a professional team sport hitherto neglected by researchers in the U.S. Using data on 361 Major League Soccer (MLS) players for one recent league season, mean and median regression models are exploited to investigate salary determinants. In comport with the available empirical evidence on racial pay discrimination in other professional team sports in the U.S., this study finds no overall evidence of pay disadvantage for non-white players. However, there is tentative evidence that black players who are not U.S. citizens actually fare worse than some other groups in salary terms.*

**JEL Codes: J31, J44, J71**

Reilly: Department of Economics, University of Sussex, Brighton, BN1 9SN, UK.
Phone: +44 (0)1273 606788; Fax: +44 (0)1273 673563; e-mail: b.m.reilly@surrey.ac.uk

Witt: Department of Economics, University of Surrey, Guildford, GU2 7XH, UK.
Phone: +44 (0)1483 686954 ; Fax: +44 (0)1483 689548 ; e-mail: r.witt@surrey.ac.uk

## I. INTRODUCTION

There exists now an extensive empirical literature on racial discrimination in professional team sports in the United States (U.S.) with perhaps the most intensively researched area focused on salary discrimination. The evidence for racial pay discrimination within professional team sports appears limited. A number of studies detected the presence of racial pay discrimination in professional basketball in the 1970s and early 1980s, but the relationship was found to have weakened substantially by the late 1990s.[1] Further, the comprehensive reviews of Khan (1991; 2000) and Rosen and Sanderson (2001) report little empirical evidence in support of racial pay discrimination in the labor markets for professional football and baseball.

Soccer is the most popular team sport in Europe and Frick (2007) notes that the market values of the largest European professional soccer clubs are comparable in magnitude to those of the most valuable sports franchises in the U.S. However, racial pay disparities in professional soccer in Europe have been the subject of minimal empirical investigation. The primary explanation for this is attributable to considerations of salary confidentiality. The limited access to information on player salaries in most European countries has forced investigators to rely on more indirect methods to inform this important policy issue.[2] For example, Szymanski (2000), using the racial composition of English league teams, found that those with an above average proportion of black players tended to perform, on average and *ceteris paribus*, better that expected given the aggregate wage bills of these clubs. The author concludes that, since players are bought and sold in a competitive labor market, this reflects evidence of market-based discrimination.[3] On the

1

other hand, Reilly and Witt (1995) and Medcalfe (2007), using data on the transfer fees that clubs pay for their players in the English league, found no evidence of unequal treatment in regard to player race, once account is taken of an array of performance and other characteristics.

Professional soccer is less popular and less well established than other major professional team sports in the U.S., and the professional game was only re-launched as recently as 1996. However, in contrast to Europe, base salary data for players are readily available, as with other professional team sports, through the relevant players' union. These data thus facilitate an analysis of pay determination of a type that is generally unfeasible for most professional European soccer leagues. The key research question motivating the current paper is whether the absence of a relationship between pay and race, as generally documented in other professional team sports, is replicated for the case of professional soccer in the U.S.

The structure of the paper is now outlined. The next section briefly sets the context for our analysis by describing the recent history, organization and nature of professional soccer in the U.S. This is followed by a section describing the dataset, and then one which reports and discusses the empirical results. A final section offers some concluding remarks.

## II. BACKGROUND

The recent history of professional soccer in the U.S. can be traced back to the formation of the North American Soccer League (NASL) in 1968. The NASL enjoyed brief success largely through the financial power of the New York Cosmos, a franchise then owned by Warner Communications, which recruited some of the world's best soccer players, though many were arguably either close to or at the end of their professional playing careers.[4] At the peak of its popularity in the late 1970s, the NASL enjoyed relatively large attendance figures and a high media profile both in the U.S. and abroad. However, many franchises acted in a financially imprudent manner and over the four year period to 1984, the league lost 17 of its franchises and ultimately collapsed in March 1985.

The successful bid by the U.S. to stage the 1994 FIFA World Cup was conditional on the establishment of a professional domestic soccer league. This occurred in 1996 with the formation of Major League Soccer (MLS), which now governs domestic professional top-flight soccer in the U.S. The financial lessons learned from the earlier NASL experience helped shape the organizational structures of MLS. The league operates as a single entity in which teams are centrally controlled and clubs are run as franchises. Revenues are shared between the league and the clubs, and player contracts are negotiated with the players' union by the league. In the 2007 season, the league comprised a total of twelve franchises spread across the U.S. and one operating in Canada. The clubs in the 2007 season were: Chicago Fire, Chivas USA, Columbus Crew, Colorado Rapids, FC Dallas, DC United, Houston Dynamo, Kansas City Wizards,

Los Angeles Galaxy, New England Revolution, New York Red Bulls, Real Salt Lake, Toronto FC.[5]

Professional soccer clubs in the U.S. are subject to a salary cap estimated to be of the order of $1.9 million for 2006. However, under a newly introduced Designated Player Rule (DPR) for the 2007 season, each MLS franchise could sign up to two players whose salaries are not governed by the cap. This ostensibly allows the league's teams to compete for selected players in the international soccer market and facilitated the much publicized arrival of David Beckham from top European club side Real Madrid to Los Angeles Galaxy.[6] Other designated players followed with the former Brazilian international Denilson moving to FC Dallas, and Colombian international striker Angel and, former United States international team captain, Reyna both signing for the New York Red Bulls.

The MLS has struggled to achieve overall profitability in its first ten years of existence. The movement of MLS clubs to specially constructed smaller soccer-specific stadia, away from the costly leasing arrangements with stadia designed primarily for use by NFL and NCAA clubs, is anticipated to enhance club and ultimately league profitability. A number of clubs who have done so (e.g., Los Angeles Galaxy and FC Dallas) are already reported to be in profit and the MLS is targeting 2010 for overall league profitability. The league is thus best viewed as one that is still navigating its way through a difficult transitional period.

### III. DATA

The data are obtained from a variety of different sources. The salary data are obtained from the MLS Players' Union and represent the base salary, which excludes bonuses and endorsements *etc*., and relate to the pay negotiated with the league prior to the start of the 2007 season. The player characteristics and performance indicators are obtained from either the official MLS website or official club websites. The racial categorization of players was determined by the inspection of color photographs, and players were allocated across three different racial groups: 'white', 'black', or 'mixed race'.[7]

The player characteristics include variables that capture age, playing position, time at the current club, whether or not the player is a U.S. citizen, whether or not the player is on the club's developmental roster, and a player's designated racial group. The performance measures include variables relating to the number of prior seasons the player has had in the MLS, the number of MLS games started in the previous season, the number of MLS games as substitute in the previous season, whether or not the player represented his country at senior international level, and whether or not the player has experience of playing abroad in higher standard European professional leagues.

The dataset used in this study comprises information on 361 professional soccer players.[8] Table 1 provides a description of the variables used in our empirical analysis and reports some selected summary statistics. It is worth noting in passing that the average player's age is 25 years, about a third of players are on a club's developmental roster revealing the youthful demographic profile of the league's players, one-third are (or were) international

players, the overwhelming majority of players are US citizens, and about two-thirds are white.  Black players represent about one-quarter of the sample, while mixed race players comprise a more modest one-tenth.  A detailed discussion of the pay data is reserved for the next section.


*Table 1 Here*

## IV. EMPIRICAL RESULTS

Table 2 provides summary statistics for the base pay measure, both overall and by race, using central tendency and dispersion statistics. The median base pay for the sample is reported at $46,331 and is modest by the standards of median earnings in other U.S. professional team sports.[9] The sample mean value of $103,297 is sensitive to the inclusion of the Beckham observation and falls by about $15,000 with the exclusion of this player's base earnings. The dispersion in pay, as measured by the Gini and the decile ratio, is high by the standards of both professional team sports and other labor markets in the U.S.[10] The table also reveals variation in average and median pay across the three broadly defined racial groups used in our analysis with black players securing lower average and median rewards compared to their white and mixed race counterparts. The ratio of median black to white base pay is 0.72 and, on average, black players lie at about the 43rd percentile ranking in the white player's pay distribution (excluding the Beckham observation). The corresponding average ranking is 48th for mixed race players and they enjoy higher median, though lower average, pay than white players.

*Table 2 Here*

Table 3 conducts statistical tests for differentials in mean and median pay outcomes across the three racial categories. The parametric F-test suggests no evidence of mean pay variation across these groups, a result which is insensitive to the exclusion of the Beckham observation. Given the reliance of the parametric test on the assumption of normality and its potential sensitivity to other outliers, the non-parametric Kruskal-Wallis

test value is also reported in this table. The test suggests, in contrast to its parametric counterpart, some weak statistical evidence of pay variation across the three racial groups. However, using a test suggested by Conover (1999, pp.218-220), there is stronger statistical evidence in support of median pay differentials across the racial groups. It should be emphasized that these test results are based on the use of raw differences in pay and do not adjust for player performance measures or other characteristics.

*Table 3 Here*

Attention now turns to table 4, which reports the mean (OLS) regression results for a number of different pay specifications. The first column of table 4 contains estimates for a conventional log pay model that includes an array of productivity characteristics, selected individual characteristics, team (or club) controls designed to capture attendance and other club specific effects like region, and two racial group variables. The majority of estimated effects for the mean regression model reported in this column are plausibly signed and well determined with the included regressors explaining about two-thirds of total variation in log base pay. An experience measure captured by the number of consecutive MLS seasons a player has accumulated enters in quadratic form and registers as a more important determinant of pay than club tenure. The point estimates suggest that base pay is maximized at about 5.7 MLS seasons. The linear term in age is also found to be important and suggests that an infinitesimal change in age raises pay by 5.9%, on average and *ceteris paribus*. A player who has acquired experience in

professional European leagues commands a sizeable average mark-up of 70% compared to a player without such experience. The number of MLS games started in the previous season yields an average *ceteris paribus* pay premium of 1.3% per additional game played, while 'bench-warmers' incur an average mark-down in pay of about 3% per substitute appearance. An international player obtains a premium of 67%, and strikers, compared to defenders and goalkeepers, enjoy a comparably sized pay premium, on average and *ceteris paribus*. Players on a club's developmental roster predictably earn less than others given their apprenticeship status, and there is no statistical evidence from this regression model that non-U.S. citizens fare less well in this labor market compared to U.S. citizens.

The point estimates for the two racial dummies are negative, though neither coefficient registers a statistically significant effect at a conventional level using two-tailed tests. The inclusion of the Beckham observation potentially overstates the magnitude of the absolute racial differential given its allocation to the white player reference group. Thus, an 'impulse' dummy netting out the Beckham observation is included in the mean regression model reported in column two of table 4. The implicit t-ratio corresponding to the dummy highlights the magnitude of the outlier effect and suggests that the Beckham observation is almost 18 standard errors above the estimated mean regression plane. However, the inclusion of the 'impulse' dummy does not alter the general narrative offered above and only marginally attenuates, in absolute terms, the estimated race effects.

The final column of table 4 reports a specification that includes a small set of race interactions which, after some experimentation, were found to be statistically important.[11] Both age and the citizenship variables were found to exert differential effects on pay for black players, while European experience was found to be important for mixed race players. In regard to black players, there does appear some evidence of unequal pay rewards in regard to age where the marginal effect is almost half that of the comparator group, which is composed of white and mixed race players. In addition, black players who are not U.S. citizens, comprising about one-half of the sub-sample of black players, incur a sizeable pay penalty compared to those black players who possess such citizenship. For instance, the average mark-down in pay for a black non-U.S. citizen of sample mean age is computed at –0.479 with a standard error of 0.137 yielding a statistically significant effect at a conventional level of significance.[12] For mixed race players the only interactive variable of importance relates to the role of European professional soccer league experience. The group of mixed race players with such experience incur a sizeable pay penalty. However, it should be stressed that this sub-group is supported by a very small cell size and some interpretational caution is thus required here.[13]

Using the relevant interactive model estimates in conjunction with the sample average for age and the sample proportion for U.S. citizenship (see table 1), the overall estimated *ceteris paribus* black pay effect can be computed. It is calculated at –0.105 with a standard error of 0.08 (prob-value = 0.189), which is close to the figure reported in column two of table 4. The corresponding estimate for the mixed race group using the

sample proportion for European experience (see table 1) is –0.116 with a standard error of 0.092 (prob-value=0.204). Thus, there is no evidence overall from the mean regression analysis of unequal treatment for either black or mixed race players in this particular labor market, but this disguises evidence of a potential unequal treatment experienced by the sub-set of black non-U.S. citizens.

*Table 4 Here*

We now briefly interrogate the variation in average base pay across the thirteen teams or franchises competing in MLS for the 2007 season. The estimates are reported in table 5 and none of the team effects was found to be individually statistically significant in the pay equations at an acceptable level nor were they found to be jointly as a set. Using the approach popularised by Krueger and Summers (1988), we report, using the final model's estimates reported in table 4, inter-team pay deviations relative to the sample weighted average. In line with the above finding, none of the estimates is statistically different from the weighted sample average and the overall variability (or dispersion) in pay across teams is fairly modest in magnitude. This clearly reflects the role of the salary cap and the nature of the tight collective bargaining arrangements governing contract base pay negotiations within this professional league.

*Table 5 Here*

An exclusive focus on the mean regression may provide a portrait of pay determination that is potentially sensitive to the role of outliers. This is likely to be an important issue in the current application given the presence within the sample of a small number of highly paid players under the PDR arrangement. The estimation of a quantile regression function at the median allows us to determine the robustness of the mean regression estimates reported in table 4. The use of such a model is also motivated by the statistically significant finding in regard to the raw median racial pay differentials reported in table 3.

In contrast to the OLS approach, the quantile regression procedure is known to be less sensitive to outliers and also provides a more robust estimator in the face of departures from normality. The former advantage is related to the fact that in the context of the Least Absolute Deviations (LAD) procedure, conventionally used in quantile regression estimation, it is the sign of the residual and not its size (as with the OLS procedure) that is important. In addition, Deaton (1997, pp.80–85) also notes that quantile regression models may possess better properties than OLS in the presence of heteroscedasticity.[14]

Using this methodology, the log pay equation can be estimated conditional on a given specification for various percentiles of the residuals (e.g., the $50^{th}$ percentile or median) by minimizing the sum of absolute deviations of the residuals from the conditional specification. The precision of the parameter estimates in a quantile regression model is dependent on the density of points at each quantile. Specifically, at those located either at the bottom or at the top end of the distribution, where the density of data points is

relatively thin, the coefficients are more difficult to compute. Furthermore, the corresponding test statistics have less statistical power, and the null hypothesis when false may not be rejected as often as it should.[15] Thus, we focus only on the estimation of the quantile regression model at the median where the density of points in the pay distribution is reasonably thick. The sampling variances for the median regression estimates are obtained in the current application using the bootstrapping procedure with 100 replications.[16]

Table 6 reports the median regression estimates for log pay specifications both with and without the set of racial interactions found to be important in the mean regression model. The estimated coefficients now provide the *ceteris paribus* effect of an explanatory variable on log pay, conditional on being at the median of the log pay distribution. In general, there appears little material difference between the mean and median regression estimates for the basic log pay specification without race interactions. However, it is noteworthy that, in contrast to the mean regression analysis, the estimated pay premium for a striker has fallen by a third perhaps suggesting that the OLS estimate reflected the inordinate influence of a small number of very highly paid strikers. In addition, the 'bench-warming' effect highlighted earlier in the mean regression discussion is now poorly determined. There is no evidence of a citizenship effect in pay and no statistical support for a *ceteris paribus* disparity in pay is uncovered in regard to race. The foregoing findings are altered little with the introduction of the race interaction variables, though there is no strong evidence now of unequal treatment with respect to age. As with the mean regression results, the statistical evidence of a pay disadvantage for the

sub-sample of black players who are not U.S. citizens remains broadly intact, though the estimated effect is now smaller in absolute terms and less well determined.[17]

The quantile regression model estimates containing the interactive variable can be used to estimate the overall pay effect for black and mixed race players. Using unconditional mean characteristics in the computation of such effects, however, may provide unrepresentative realizations for these characteristics at points other than the conditional mean pay to which they relate. Therefore, it is necessary to use realizations of the characteristics that more accurately reflect the relevant point on the conditional pay distribution, which is the median in this case. In order to address this issue, we use an approach originally suggested by Machado and Mata (2005) to derive the realizations for the relevant characteristics at the median of the conditional pay distribution. The procedure involves drawing 100 observations at random and with replacement from the original sample of 361. Each observation once ranked comprises a percentile point on the log pay distribution. The relevant set of characteristics for the observation at the median is then retrieved, which in the current case is comprised of a player's age, U.S. citizenship and European experience. This process is then replicated 300 times to obtain 300 median observations on these three variables. The mean values of these median observations can then be used to construct overall pay effects by race for the median regression.[18]

Using these conditional mean values, the overall estimated *ceteris paribus* black pay effect is now computed at –0.045 with a standard error of 0.097 (prob-value = 0.64), and

the overall estimate for the mixed race group is 0.017 with a standard error of 0.122 (prob-value=0.88). The quantile estimates serve to confirm in regard to race the broad tenor of our overall findings from the mean regression analysis. We can also conclude that once productivity and other measures are controlled for, the statistically significant median raw pay differences by race, noted in table 3, disappear. However, the estimated interactive effects suggest that there remains some very tentative evidence of a pay disadvantage for those black players who are not U.S. citizens (see footnote 17).

*Table 6 Here*

# V. CONCLUSIONS

The availability of salaries for MLS players for the 2007 season has allowed the seminal investigation of pay determination for a professional team sport that has been neglected by researchers in the U.S. to date. The lack of an apparent research interest in this sport's labor market is understandable given its size and market valuation compared to football, baseball or basketball. A casual comparison of median base pay in soccer with the remuneration available in the more established professional team sports confirms the 'Cinderella' status of professional soccer within the broader U.S. sports industry. Nevertheless, our empirical analysis of the pay determination process revealed that, in comport with the analyses of salary within other labor markets for professional team sports, a sizeable part of variation in base salary was found to be explained by the variation in player characteristics and performance measures.

The paper finds no statistical evidence of variation in pay across the teams that compete in the MLS. This is perhaps to be anticipated given the existence of a salary cap and the fact that base pay contracts are negotiated between the league and the players' union within a collective bargaining framework. The MLS, in contrast to its progenitor, is apparently doing an effective job in ensuring a level playing field in terms of intra-league pay differentials. The greater part of base pay dispersion in the MLS appears to be intra-team rather than inter-team in nature.

Our key result on race is resonant of the findings for most other professional team sports in the U.S. but it is somewhat nuanced. On balance, we found no persuasive statistical

evidence of an unequal treatment by racial group using mean regression analysis regardless of whether or not a large and well documented outlier was netted out of the analysis. In addition, the use of median regression analysis confirmed the robustness of our key results. However, there was some evidence that the possession of U.S. citizenship appeared a more favorable asset for black players compared to all other groups and this manifested itself in the form of an unequal treatment for those black players who do not have the benefit of such citizenship.

Future analysis is clearly merited to determine whether these findings are robust to the conduct of a richer empirical analysis perhaps incorporating more detailed performance indicators than currently available. In addition, it will also be of some interest to ascertain whether our findings on race represent a transient or a more persistent feature of the pay structure within a labor market that is an integral part of a professional sport that currently operates within an uncertain and transitional environment.

# REFERENCES

Brown, E., Spiro, R., and D.Keenan. "Wage and Non-wage Discrimination in Professional Basketball: Do Fans Affect It?" *American Journal of Economics and Sociology*, 50(3), 1991, 333-345.

Brownstone, D. and Valletta, R. "The Bootstrap and Multiple Imputations: Harnessing Increased Computing Power for Improved Statistical Tests", *Journal of Economic Perspectives*, 15(4), Fall 2001, 129-141.

Conover, W.J. *Practical Nonparametric Statistics*, Third Edition, Wiley Series in Probability and Statistics, 1999.

Cowell, F.A., "Sampling Variance and Decomposable Inequality Measures", *Journal of Econometrics*, 42, 1989, 27–41.

Deaton, A. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Baltimore: John Hopkins University Press, 1997.

Dey, M. "Racial Differences in National Basketball Association Players' Salaries: Another Look." *American Economic Review*, 41(2), 1997, 84-90.

Frick, B. "The Football Players' Labor Market: Empirical Evidence from the Major European Leagues." *Scottish Journal of Political Economy*, 54(3), 2007, 422-446.

Hamilton, B.H. "Racial Discrimination and professional Basketball Salaries in the 1990s."*Applied Economics*, 29, 1997, 287-296.

Khan, L.M. "Discrimination in Professional Sports: A Survey of the Literature." *Industrial and Labor Relations Review*, 44(3), 1991a, 395-418.

Khan, L.M. "The Sports Business as a Labor Market Laboratory." *Journal of Economic Perspectives*, 14(3), Summer 2000, 75-94.

Khan, L.M. and P.D.Sherer. "Racial Differences in Professional Basketball Players' Compensation." *Journal of Labor Economics*, 6(1), 1988, 40-61.

Koenker, R. *Quantile Regression*, Econometric Society Monographs, Cambridge University Press, 2005.

Koenker, R. and K.F.Hallock "Quantile Regression", *Journal of Economic Perspectives*, 15(4), Fall 2001, 143-156.

Koenker, R. and G. Bassett. "Regression Quantiles", *Econometrica*, 46, 1978, 33-50.

Krueger, A. and L. Summers. "Efficiency Wages and the Inter-industry Wage Structure", 1988, *Econometrica*, 56(2), 259-293.

Lucifora, C. and R.Simmons. "Superstar Effects in Sports: Evidence from Italian Soccer." *Journal of Sports Economics*, 4, 2003, 35-55.

Machado, J. and Mata, J. "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression Models", *Journal of Applied Econometrics*, 20,2005, pp. 445-465.

Medcalfe, S. "English League Transfer Prices: Is There a Racial Dimension? A Re-examination with New Data." *Applied Economics Letters*, forthcoming.

Preston, I and S.Szymanski. "Racial Discrimination in English Football". *Scottish Journal of Political Economy*, 47(4), 2000, 342-363.

Reilly, B. and R.Witt. "English League Transfer Prices: Is There a Racial Dimension?" *Applied Economics Letters*, 2, 1995, 220-222.

Rosen, S. and A. Sanderson. "Labour Markets in Professional Sports." *The Economic Journal*, 111, February 2001, F47-F68.

Szymanski, S. "A Market Test for Discrimination in the English Professional Soccer Leagues." *Journal of Political Economy*, 108(3), 2000, 590-603.

**Footnotes:**

1. Kahn and Sherer (1988) and Brown, Spiro and Keenan (1991) provide some evidence on this topic for the early period with Hamilton (1997) and Dey (1997) reporting evidence for the later era.
2. Lucifora and Simmons (2003) provide a notable exception in this context with their analysis of player earnings in the top two professional tiers of the Italian soccer league.
3. Subsequent research undertaken by Preston and Szymanski (2000) eliminated fan (or consumer) motivated behaviour as a potential source for the discrimination phenomenon detected.
4. The players enlisted by the Cosmos included FIFA World Cup winning captains Beckenbauer (for West Germany in 1974) and Carlos Alberto (for Brazil in 1970). However, the most significant signing, lured out of semi-retirement in 1975 by a lucrative three-year contract reputed to be close to $7 million (at 1975 prices), was the Brazilian Pelé, universally acknowledged as the world's greatest ever soccer player.
5. The teams are divided into the Western Conference composed of Chivas USA, Houston Dynamos, FC Dallas, Colardo Rapids, Los Angeles Galaxy, and Real Salt Lake, and the Eastern Conference comprising the seven remaining teams.
6. Beckham signed a five-year contract with Los Angeles Galaxy. The annual base pay deal guaranteed $5.5 million per year but with endorsements the deal could potentially yield close to $200 million over the lifetime of the contract. This would make it the most lucrative contract ever agreed for a professional soccer player.
7. For the overall majority of players, no ambiguous cases were encountered in their allocation to racial groups. In the small number of cases where such ambiguity arose, the allocation was informed by the advice of a team of graduate students from a variety of different racial backgrounds. The full allocation of players to the three racial groups is available from the corresponding author on request.
8. This compares favorably to the sample sizes generally available in studies of this kind undertaken for other professional team sports (see Khan (1991)).
9. In order to situate these pay data for professional soccer within a broader context, median base annual earnings in the U.S in 2007 was approximately $2.6 million, $1 million and $0.6 million for professional players in basketball, baseball and football respectively.
10. The magnitude of pay inequality is not surprising in this labor market given that the data comprise players at both extremes of the profession: those securing 'superstar' levels of remuneration and those on the developmental rosters whose base pay is close to statutory minimum wage levels. For instance, some players on the developmental rosters earn an annual base pay of $12,900.
11. A 'general-to-specific' modelling approach was adopted commencing with a fully interactive model comprising a set of black and mixed race interactions for all variables other than the club controls. This was sequentially tested down in order of statistical insignificance, using a tight criterion for individual statistical

significance, to the more parsimonious model reported in the final column of table 4.

12. The estimates actually suggest that black U.S. citizens fare better than all other groups in the sample.

13. There are only five mixed race players with European club experience out of the sample of 361 data points used in our analysis.

14. See Koenker (2005) and Koenker and Basset (1978) for a detailed discussion of quantile regression models, and Koenker (2001) for an accessible introduction to the topic.

15. Hamilton (1997), with fewer data points on professional basketball salaries than available in the current study, elected to estimate quantile regression models at the $10^{th}$ and $90^{th}$ percentiles of the conditional pay distribution. This is not an approach considered prudent to pursue here.

16. See Brownstone and Valletta (2001) for a simple discussion of bootstrapping methods in applied econometrics.

17. Although neither the 'black' nor 'age' estimates are well determined in the median regression model, the estimated pay penalty for a black non-U.S. citizen with a conditional average age consistent with being at the median of the log pay distribution (see footnote 18) is computed at –0.35 with a standard error of 0.18, implying a prob-value of 0.053. Given that the result's statistical significance is largely being driven by the sample covariances between the estimates and not their sampling variances, this finding should be interpreted tentatively and with some degree of caution.

18. The average and sample proportion values of the three characteristics, conditional on being at the median of the conditional log pay distribution, are 23.64, 0.79 and 0.08 for age, US citizenship and European experience respectively.

**TABLE 1**

**Variable Description and Summary Statistics**

| Variable | Variable Description | Mean |
|---|---|---|
| **Log(Pay)** | The natural logarithm of the base annual salary for the 2007 season (source: MLS Players' Union). | 10.813 (1.022) |
| **Age** | Player's age measured in years. | 25.086 (4.303) |
| **Seasons** | The number of consecutive seasons the player has been in the MLS to the start of the 2007 season. | 2.416 (2.948) |
| **Tenure** | The number of months the player has been at the current club to the start of the 2007 season. | 15.316 (20.765) |
| **Games Started** | The number of games started in the MLS in the previous (2006) season. | 9.925 (11.448) |
| **Games as Substitute** | The number of games played as a substitute in the MLS in the previous (2006) season. | 2.061 (3.420) |
| **International** | = 1, if the player is an international and = 0, otherwise. | 0.319 |
| **Striker** | = 1, if the player is a forward and = 0, otherwise. | 0.227 |
| **Midfield** | = 1, if the player is a midfielder and = 0, otherwise. | 0.380 |
| **Other**[§] | = 1, if the player is either a defender or a goalkeeper and = 0, otherwise. | 0.393 |
| **Development** | = 1, if the payer is on the club's developmental roster and = 0, otherwise. | 0.305 |
| **White**[§] | = 1, if the player is white and = 0, otherwise. | 0.654 |
| **Black** | = 1, if the player is black and = 0, otherwise. | 0.241 |
| **Mixed** | = 1, if the player is mixed race and = 0, otherwise. | 0.105 |
| **Citizen** | = 1, if the player is a US citizen and =0, otherwise. | 0.698 |
| **European** | = 1, if the player has had experience of playing in European professional leagues and =0, otherwise. | 0.163 |

Notes to table 1:

(a) Summary statistics are computed using 361 observations;

(b) Standard deviations are reported in parentheses for continuous variables only;

(c) § denotes omitted category in econometric estimation.

<div align="center">**TABLE 2**</div>

<div align="center">**Summary Basic Pay Statistics for MSL 2007 Season**</div>

| | All Players | All Players (excl. Beckham) | White Players | White Players (excl. Beckham) | Black Players | Mixed Race Players |
|---|---|---|---|---|---|---|
| **Mean** | 103,297 | 88,306 | 119,409 | 96,513 | 66,996 | 86,342 |
| **Median** | 46,331 | 46,116 | 50,200 | 50,000 | 36,000 | 60,000 |
| **Standard Deviation** | 335,815 | 178,141 | 409,206 | 209,565 | 97,072 | 82,207 |
| **Gini Coefficient** | 0.6284 (0.0261) | 0.5689 (0.0184) | 0.6574 (0.0286) | 0.5820 (0.0227) | 0.5382 (0.0220) | 0.4843 (0.0492) |
| **$90^{th}/10^{th}$ Percentile Ratio** | 14.6 | 13.8 | 13.95 | 13.95 | 12.09 | 15.74 |
| **Sample Size** | 361 | 360 | 236 | 235 | 87 | 38 |

Notes to table 2:

(a) Summary statistics for pay measure reported here are based on its non-logged form;

(b) Values reported for the mean, median and standard deviation are expressed in US dollars;

(c) Standard errors reported for the Gini coefficient are computed using the INEQ programme (see Cowell (1989))

<div align="center">**TABLE 3**</div>

**Statistical Tests for Base Pay Differences by Race**

| Test Statistic | Pay Differences by Race |
|---|---|
| F-test (2,358) | 0.83 (0.438) |
| F-test (2,357) (excl. Beckham) | 0.87 (0.418) |
| Kruskal-Wallis Test | 4.625* (0.097) |
| Median Test | 6.869** (0.032) |

Notes to table 3:

(a) F-tests are based on one-way ANOVA.

(b) The Kruskall-Wallis test is distributed as a chi-squared with two degrees of freedom.

(c) Median test is computed using Conover (1999, pp.218-220) and is distributed as a chi-squared with two degrees of freedom.

(d) ***, ** and * denote statistical significance at the 0.01, 0.05 and 0.1 level respectively.

(e) Prob-values are reported in parentheses.

TABLE 4
**Mean Regression Estimates of Base Pay Determinants for MLS Players**

| Variable | Model (1) | Model (2) | Model (3) |
|---|---|---|---|
| **Constant** | 8.9238*** | 9.0400*** | 9.0010*** |
| | (0.3628) | (0.3462) | (0.3974) |
| **Age** | 0.0590*** | 0.0531*** | 0.0616*** |
| | (0.0148) | (0.0139) | (0.0150) |
| **Seasons** | 0.1763*** | 0.1918*** | 0.1899*** |
| | (0.0502) | (0.0471) | (0.0447) |
| **Seasons$^2$** | -0.0155*** | -0.0161*** | -0.0156*** |
| | (0.0043) | (0.0041) | (0.0040) |
| **Tenure** | -0.0008 | -0.0005 | 0.0003 |
| | (0.0018) | (0.0016) | (0.0016) |
| **Games Started** | 0.0129*** | 0.0132*** | 0.0114*** |
| | (0.0038) | (0.0037) | (0.0036) |
| **Games as Substitute** | -0.0283*** | -0.0290*** | -0.0237*** |
| | (0.0103) | (0.0101) | (0.0094) |
| **International** | 0.5132*** | 0.4806*** | 0.5187*** |
| | (0.0878) | (0.0826) | (0.0816) |
| **Striker** | 0.4626*** | 0.4662*** | 0.4433*** |
| | (0.0972) | (0.0975) | (0.0923) |
| **Midfield** | 0.2828*** | 0.2500*** | 0.1929*** |
| | (0.0754) | (0.0696) | (0.0678) |
| **Development** | -0.4787*** | -0.3725*** | -0.3707*** |
| | (0.0870) | (0.0836) | (0.0840) |
| **Citizen** | -0.0908 | -0.0756 | -0.2856*** |
| | (0.0851) | (0.0841) | (0.1108) |
| **European** | 0.5332*** | 0.5221*** | 0.5162*** |
| | (0.1067) | (0.1061) | (0.1095) |
| **Black** | -0.1311 | -0.1105 | 0.5006 |
| | (0.0841) | (0.0813) | (0.4592) |
| **Mixed** | -0.0793 | -0.0523 | -0.0198 |
| | (0.1024) | (0.0991) | (0.0980) |
| **'Beckham'** | † | 3.6387*** | 3.4203*** |
| | | (0.2041) | (0.2150) |
| **Black×Age** | † | † | -0.0392** |
| | | | (0.0182) |
| **Black×Citizen** | † | † | 0.5420*** |
| | | | (0.1674) |
| **Mixed×European** | † | † | -0.5935*** |
| | | | (0.2027) |
| **Team Controls** | Twelve Included | Twelve Included | Twelve Included |
| R$^2$ | 0.6773 | 0.7097 | 0.7304 |
| Adjusted-R$^2$ | 0.6522 | 0.6862 | 0.7059 |
| Breusch-Pagan Test for Homoscedasticity | 156.3 (0.000) | 109.2 (0.000) | 109.3 (0.000) |
| Sample Size | 361 | 361 | 361 |

Notes to table 4:

(a)  The dependent variable is the natural logarithm of a player's base salary (see table 1 for definition).

(b)  Robust standard errors are reported in parentheses based on Huber-White given the presence of heteroscedasticity (see Breusch-Pagan test result).

(c) ***, ** and * denote statistical significance at the 0.01, 0.05 and 0.1 level respectively using two-tailed tests.

(d) 'Beckham' is an impulse dummy taking a value of 1 for the player's observation and 0 otherwise.

(e)  The team controls are Chicago Fire, Chivas USA, Columbus Crew, Colorado Rapids, FC Dallas, DC United, Houston Dynamo, Kansas City Wizards, Los Angeles Galaxy, New England Revolution, New York Red Bulls, Real Salt Lake, Toronto FC (with Chicago Fire providing the reference group in estimation).

(f) The prob-value is reported in parentheses for the Breusch-Pagan test.

(g) † denotes not applicable in estimation.

**TABLE 5**

**Inter-team Base Pay Differentials**

| Team | Deviation from Average |
|---|---|
| **Chicago Fire** | 0.1067 |
| | (0.1031) |
| **Chivas USA** | -0.0805 |
| | (0.0791) |
| **Columbus Crew** | -0.0038 |
| | (0.0872) |
| **Colorado Rapids** | -0.0511 |
| | (0.0911) |
| **FC Dallas** | 0.1215 |
| | (0.0922) |
| **DC United** | -0.0161 |
| | (0.1164) |
| **Houston Dynamo** | -0.0853 |
| | (0.1017) |
| **Kansas City Wizards** | 0.0471 |
| | (0.0888) |
| **Los Angeles Galaxy** | 0.0075 |
| | (0.1152) |
| **New England Revolution** | -0.0852 |
| | (0.0851) |
| **New York Red Bulls** | -0.0257 |
| | (0.1222) |
| **Real Salt Lake** | 0.1150 |
| | (0.1226) |
| **Toronto FC** | -0.0552 |
| | (0.1011) |
| Overall Variability | 0.073 |
| Wald Test for Overall Significance of Team Effects | 6.720 |
| | (0.876) |

Notes to table 5:

(a) Deviations from the average are based on Krueger and Summers (1988), as is the overall variability measure reported, and includes the Beckham 'impulse' dummy.

(b) Robust standard errors are reported in parentheses based on Huber-White.

(c) ***, ** and * denote statistical significance at the 0.01, 0.05 and 0.1 level respectively using two-tailed tests.

(d) The adjusted deviations are based on regression model estimates reported in column three of table 4 (i.e., Model (3)).

(e) The Wald test is computed using the heteroscedasticity corrected variance-covariance matrix using the regression model reported in column three of table 4. The test statistic is distributed with 12 degrees of freedom and the prob-value is reported in parentheses.

TABLE 6

**Median Regression Estimates of Base Pay Determinants for MLS Players**

| Variable | Model (1) | Model (2) |
|---|---|---|
| **Constant** | 8.9363*** | 9.2218*** |
| | (0.4667) | (0.4917) |
| **Age** | 0.0527*** | 0.0501** |
| | (0.0178) | (0.0171) |
| **Seasons** | 0.2195*** | 0.1928*** |
| | (0.0621) | (0.0584) |
| **Seasons$^2$** | -0.0177*** | -0.0149** |
| | (0.0051) | (0.0061) |
| **Tenure** | -0.0015 | -0.0013 |
| | (0.0020) | (0.0020) |
| **Games Started** | 0.0149*** | 0.0147*** |
| | (0.0044) | (0.0050) |
| **Games as Substitute** | -0.0121 | -0.0168 |
| | (0.0116) | (0.0124) |
| **International** | 0.4363** | 0.4394*** |
| | (0.0130) | (0.0995) |
| **Striker** | 0.2845*** | 0.3047*** |
| | (0.0900) | (0.0948) |
| **Midfield** | 0.1378* | 0.1486** |
| | (0.0723) | (0.0721) |
| **Development** | -0.5038*** | -0.5656*** |
| | (0.1163) | (0.1193) |
| **Citizen** | -0.1217 | -0.2806** |
| | (0.1099) | (0.1422) |
| **European** | 0.5152*** | 0.4985** |
| | (0.1182) | (0.1320) |
| **Black** | -0.0959 | 0.6076 |
| | (0.0929) | (0.6548) |
| **Mixed** | 0.0323 | 0.0733 |
| | (0.1223) | (0.1266) |
| **Black×Age** | † | -0.0405 |
| | | (0.0282) |
| **Black×Citizen** | † | 0.3856** |
| | | (0.1802) |
| **Mixed×European** | † | -0.7014** |
| | | (0.2999) |
| **Team Controls** | Twelve Included | Twelve Included |
| **Pseudo-R$^2$** | 0.518 | 0.534 |
| **Sample Size** | 361 | 361 |

Notes to table 6:
(a) The dependent variable is the natural logarithm of a player's base salary (see table 1 for definition).
(b) Standard errors are reported in parentheses and are based on the bootstrapping method with 100 replications.
(c) ***, ** and * denote statistical significance at the 0.01, 0.05 and 0.1 level respectively using two-tailed tests.
(d) The team controls used are those listed in table 4.